Pimentez vos week-end !

QCM

MODE D'EMPLOI

S.Puget
DIU de pédagogie médicale
2010

A mettre entre (presque) toutes les mains !

Introduction

La naissance des QCM daterait d'il y a à peu près 80 ans et est attribuée à un psychométricien américain, A.C Eurich (Eurich) qui proposait de remplacer dans certains cas les questions rédactionnelles (ouvertes) par des questions à réponse fermée. Les QCM sont un outil d'évaluation qui peut être utilisé pour une évaluation diagnostique (en amont d'une formation), formative (en cours de formation) ou sommative (en fin de formation).

Ils sont maintenant largement utilisés dans l'enseignement supérieur essentiellement pour des évaluations sommatives. Pourtant, leur réputation auprès des enseignants est mitigée et ils les considèrent par nature comme inférieurs aux questions ouvertes. Les étudiants qui rencontrent des QCM mal formulés les considèrent comme déroutant et réducteurs. Depuis l'apparition des questions à réponse fermée dont les QCM sont le modèle de base, on assiste en effet au débat des partisans des questions à réponse ouverte versus fermée dont les avantages et inconvénients sont souvent opposés.

Beaucoup d'enseignants pensent que les QCM sont adaptés pour tester des informations factuelles alors que les questions ouvertes testeraient des connaissances d'un plus haut niveau cognitif. Cette critique à l'encontre des QCM est plus souvent à attribuer à un défaut de construction des items qui le constituent plutôt qu'à leur propre faiblesse (Collins, McCoubrie).

Les QCM peuvent tester beaucoup de connaissances en peu de temps, de même niveau cognitif que les questions rédactionnelles si ils sont écrits selon certains critères qui en feront un outil d'évaluation des connaissances fiable et valide (Boland, Braddom, Case, Collins, Cox, CTL, Farley, Gronlund, Haladyna, Norcini Pagonis, Schulteis, Vydareny).

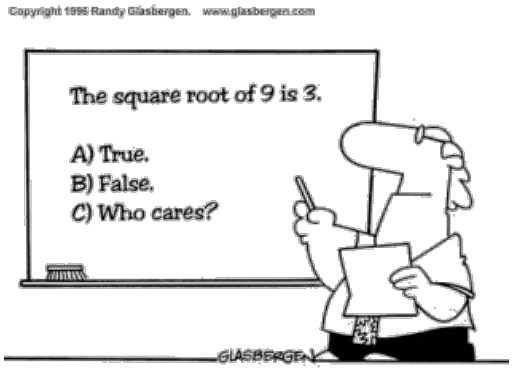
Nous proposons dans ce mémoire de résumer les bonnes pratiques de rédaction des QCM formulées dans la littérature. Quatre points seront abordés : 1) rédiger les objectifs éducationnels ; 2) définir les niveaux de connaissance pour chaque objectif, 3) rédiger des QCM appropriés pour tester ces connaissances et 4) analyser les réponses.

Rédiger les objectifs Educationnels

L'écriture d'un bon test de connaissance doit commencer par l'identification des informations ou connaissances à tester. Il doit y avoir une relation directe entre les objectifs d'éducation et les items du test. Ainsi, les items du test doivent découler des objectifs (Kemp) et se focaliser sur les questions de fond importantes en évitant de tester des connaissances médicales triviales. Les items controversés doivent être proscrits (Braddom). La détermination des questions appropriées pour le test peut être facilitée par la revue des points importants et l'identification de phrases qui résument les idées principales du sujet (Collins). A partir de là, les points clefs peuvent être écrits sous forme de phrases simples, à l'origine d'un schéma clair des connaissances qui doivent être acquises par l'étudiant.

La rédaction des objectifs doit tenir compte du comportement de l'apprenant et pas seulement du programme lui même. Les objectifs doivent définir les connaissances les plus importantes à acquérir. Les objectifs « mesurables » permettent d'estimer le mieux si un étudiant a acquis un objectif. Exemples de termes mesurables « spécifier, expliquer, lister, identifier, comparer… » Et de termes non mesurables « savoir, comprendre, apprendre, être familier avec….». Par exemple, *un objectif non mesurable* « connaitre l'aspect d'un hématome extra dural sur un scanner cérébral » (on ne sait pas clairement comment l'étudiant va pouvoir montrer qu'il connaît) ; *un objectif mesurable* « décrire 3 caractéristiques d'un hématome extra dural sur un scanner cérébral » (on voit plus clairement comment l'étudiant va exposer ses connaissances, le terme « 3 caractéristiques » indiquant un niveau de connaissance spécifique) (Collins).

L'utilisation des QCM doit être adaptée aux objectifs que l'on souhaite tester (annexe 1).



Many students actually look forward to Mr. Atwadder's math tests.

Définir les niveaux cognitifs

Les enseignants savent qu'il existe plusieurs niveaux de connaissance mais manquent parfois d'un système simple pour appliquer ce concept dans la préparation d'un cours ou la réalisation d'un examen (Collins, CTL). En 1956, Bloom publie une taxonomie des objectifs pédagogiques prenant en compte connaissance, compréhension, application, analyse, synthèse et évaluation (Figure 1 tirée de Woodford) (Bloom).

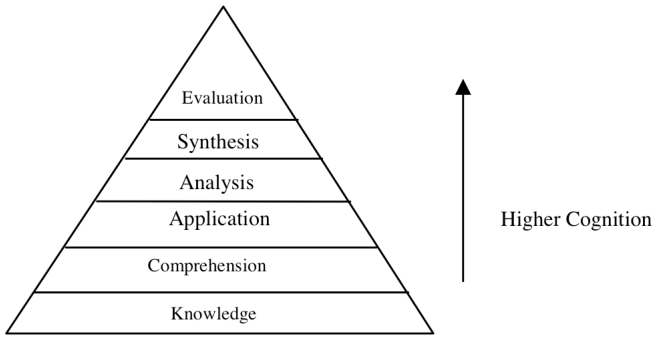


Figure 1: Bloom's levels of cognition

Depuis, un système plus simple à 3 niveaux a été proposé (Crooks):

-Niveau 1 : Connaissance (*Rappel ou reconnaissance)*. L'étudiant se souvient de faits, de principes ou de théories mémorisés.

-Niveau 2 : *Application et compréhension*. L'étudiant est capable de comprendre et d'expliquer ce qu'il a appris avec ses propres mots, et utiliser de nouvelles informations, théories ou principes. Il doit translater ses connaissances pour résoudre un nouveau problème.

-Niveau 3: *Analyse, Résolution de problèmes.* L'étudiant doit examiner une information, identifier des causes, des motifs, faire des déductions, trouver la structure sous-jacente et les relations.

La plupart des QCM utilisés ne teste que le niveau 1. Cependant, faire ses cours en tenant compte de ces 3 catégories peut aider à préparer des questions qui testent à la fois l'étendue des connaissances et les différents niveaux cognitifs auxquels l'étudiant doit faire appel. Cette démarche paraît essentielle pour formuler des QCM à haut niveau de validité et fiabilité. Le QCM doit tester les connaissances de même niveau que l'objectif qui est évalué (Table 1 : exemple de QCM et objectif pour chaque niveau de connaissance).

Validité et Fiabilité

Les 2 plus importantes caractéristiques d'un test de connaissances sont sa validité et sa fiabilité. *La validité* d'un test est sa capacité à tester les connaissances et compétences que l'étudiant était sensé acquérir durant la formation. *La fiabilité* est la capacité à discriminer de façon consistante des étudiants de niveaux différents.

Un QCM bien rédigé est en général plus valide et fiable qu'une question rédactionnelle car (1) il teste des connaissances plus larges ; (2) la discrimination entre les différents niveaux de connaissances est plus facile à déterminer et (3) la cohérence de la notation est en théorie garantie (CTL).

La validité d'un QCM est basée sur une sélection systématique des items en prenant en compte à la fois le contenu (fond ou étendue des connaissances) et le niveau (cognitif) des connaissances. Les enseignants ont plutôt tendance à tester le contenu et à négliger le niveau de la question choisie. De plus les QCM qui ne font appel qu'à la mémoire ou à la reconnaissance sont les fréquents car les plus faciles à rédiger.

Guide de rédaction des QCM

Les différents composants d'un QCM ont des noms particuliers : Un « item » constitue une unité composé d'un énoncé et des différentes options de réponse. L'énoncé (amorce ou tronc) est une question ou une déclaration. Les réponses sont appelées « alternatives », « options » ou « choix ». La bonne option est appelée « réponse correcte » et les options incorrectes « distracteurs ». Il existe plusieurs types de QCM dont les plus fréquents sont données en exemple dans l'annexe 2.

D'après le guide de rédaction du Center Teaching and Learning (CTL), il est recommandé d'écrire quelques questions après chaque cours, quand tout est encore fraichement inscrit dans votre mémoire ; l'examen sera ainsi un meilleur reflet des points (acquis) sur lesquels vous avez insisté. Le principe de base de l'écriture d'un QCM est relativement simple : la question doit être formulée de façon à ce que les étudiants doués pour les tests ne soient pas favorisés et que ceux qui le sont moins ne soient pas pénalisés. Le guide de rédaction qui suit est fait pour écrire des QCM qui mesure les connaissances plutôt que les compétences pour répondre à des tests.

Généralités

- Construire chaque item pour ne répondre qu'à un seul objectif bien identifié
- Chaque question doit se suffire à elle même et les items doivent être indépendants. Eviter de poser des items qui dépendent de connaître les réponses aux autres questions. Vérifier que l'information donnée par certains items ne donne pas des indices pour répondre à d'autres.
- Eviter les termes absolus comme « toujours », « jamais », « tous » , « systématiquement » ou « aucun » dans l'énoncé et les réponses.
- Eviter les termes imprécis comme « peut », « pourrait » et les termes « rarement », « parfois », « occasionnellement », « peu », « beaucoup ».
- Eviter les éponymes, acronymes et abréviations sans les renseigner. Les étudiants peuvent ne pas être familiers avec certains termes. De plus, ce genre de terme peut avoir plus d'une signification.
- La fiabilité ou pouvoir discriminant d'un QCM est basé sur la qualité et l'attractivité des distracteurs.

Ecrire l'Enoncé

L'énoncé d'un QCM pose un problème ou formule une question. Il peut comporter un mot, une phrase, une courte histoire clinique ou une iconographie. La règle de base est que l'étudiant doit être capable de comprendre la question dès la première lecture et sans avoir à lire les différentes propositions de réponse.

- La qualité des consignes conditionne la validité des résultats : L'énoncé type de QCM doit être précisé (préciser 1 ou plusieurs réponses correctes…).

- Le problème posé doit être simple et unique, les questions directes étant les meilleures. L'énoncé peut comprendre un graphe, un diagramme ou des examens radiologiques mais doit toujours être accompagné d'une phrase complète plutôt que d'une proposition à compléter (annexe 3). Il a été montré en effet qu'une phrase incomplète diminue le taux de bonne réponse des élèves de 10 à 15% (Kent). Parfois un énoncé incomplet peut être nécessaire pour éviter un langage alambiqué ou des phrases trop compliquées mais cela doit rester exceptionnel.

- L'énoncé doit être bref, évitant les propos prolixes et doit contenir toute l'information et seulement l'information pertinente pour pouvoir répondre correctement à la question (annexe 4). Il convient de veiller à : l'univocité (le mot garde le même sens dans des emplois différents), la précision (le terme ne laisse place à aucune indécision), la simplicité (le niveau de langage correspond à celui des élèves) et l'honnêteté (on évitera d'introduire des notions non connues) (Bonnefon). On considère qu'un étudiant doit pouvoir répondre à 1 à 2 QCM par minute (Lowman, Srinivasa).

  Tout en restant bref, les questions testant un plus haut niveau cognitif sont souvent plus longues que les autres. Rédiger des QCM qui mesurent les hauts niveaux cognitifs est difficile à mettre en place pour de nombreux élèves sans y consacrer beaucoup de temps. Dans de telles questions d'évaluation, une situation est décrite dans un court paragraphe et le problème est posé comme énoncé de la question. Toutes les règles de rédaction des QCM sont applicables mais les étudiants doivent faire appel à leur jugement et à une réflexion critique pour y répondre correctement (annexe 5).

- L'énoncé doit être posé sous forme positive, les étudiants pouvant mal interpréter des phrases négatives qui d'ailleurs conduisent à des QCM moins valides (Braddom). Les questions négatives comportent souvent « laquelle n'est

pas vraie… » ou « toutes sauf… ». Ces formulations négatives doivent être proscrite sauf cas exceptionnel. Si elles sont utilisées, les propos négatifs doivent être soulignés ou inscrits en lettres capitales ou en italique (annexe 6). Il faut d'autant plus proscrire les doubles négations.

- L'énoncé ne doit pas introduire pas de jugement de valeur (pas de données raciales ou de genre. Par exemple : éviter de ne donner que des exemples de femmes dépressives).

Ecrire les Réponses

En principe, un QCM ne doit pas comporter moins de 3 options de façon à ce que les étudiants puissent le moins facilement deviner la ou les bonne(s) réponse(s). Le nombre optimal entre 3 et 5 options est débattu (Haladyna, Woodford). Ne choisir préférentiellement qu'une seule bonne réponse. Si il y a plusieurs bonnes réponses cela peut devenir un jeu frustrant pour l'etudiant qui va essayer de deviner ce que l'enseignant avait en tête en écrivant l'item (Burton). Les règles de base pour écrire les questions sont (1) l'étudiant doit pouvoir sélectionner le bon item sans avoir à s'occuper de complexités qui n'ont rien à voir avec la connaissance de la bonne réponse (2) l'étudiant ne doit pas pouvoir deviner la bonne réponse en fonction de la façon dont elle est rédigée.

- Ecrire la bonne réponse en premier après avoir écrit l 'énoncé en s'assurant qu'il s'agit <u>incontestablement de la bonne réponse</u>. Dans le cas où il s'agirait de « la meilleure réponse possible », s'assurer que des experts seraient d'accord pour que ce soit la meilleure réponse possible.

- <u>Ecrire les distracteurs de la même façon que la réponse correcte</u> en <u>longueur, complexité, formulation, style et contenu (annexe 7)</u>. L'écriture de distracteurs plausibles est un exercice difficile. Le meilleur distracteur est une déclaration exacte mais qui ne remplit pas toutes les conditions du problème ou une déclaration incorrecte qui paraît juste pour l'étudiant (Davis). On peut augmenter la vraisemblance d'un distracteur en y incluant des informations superflues ou en basant les distracteurs sur des erreurs logiques ou courantes, en évitant d'utiliser des termes non familiers pour les étudiants (Collins).

- Chaque proposition ne doit contenir <u>qu'une notion</u>

- Deux propositions ne peuvent ni être mutuellement exclusives ni interdépendantes l'une de l'autre

- Eviter de composer des alternatives qui ne comportent que de minimes distinctions entre elles, à moins que la capacité à les distinguer soit un objectif particulier formulé dans le cours. Eviter également le chevauchement de termes (annexe 8).

- Eviter d'écrire « toutes les autres réponses», favorisant les étudiants qui n'ont qu'une connaissance partielle à deviner la bonne réponse, pour les bons, quant à eux, il suffit de trouver 2 réponses correctes pour savoir que c'est la bonne réponse. Eviter d'écrire « aucune des autres réponses ». Cela ramène l'item à un vrai/faux et n'est vraiment approprié que pour les matières pour lesquelles il existe des réponses correctes absolues comme en mathématiques (annexe 9).

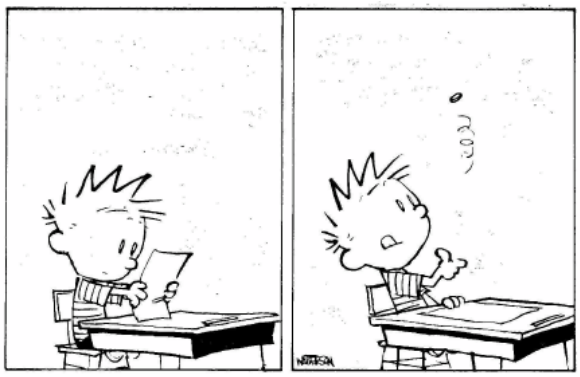
- Eviter de donner des indices de langage qui donnent la bonne réponse. Cela inclus les erreurs de grammaire ou de syntaxe, des mots clefs qui n'apparaissent que dans la bonne réponse. Cela n'est pas un problème quand la question est écrite sous forme de question (annexe 10). Eviter également d'écrire la réponse correcte dans un langage parfait et les distracteurs dans un langage familier ou d'écrire deux réponses différentes qui ont la même signification.

- Randomiser la position des bonnes réponses. On peut par exemple ranger les réponses par ordre alphabétique ou bien utiliser des tables de randomisation. Exemple ci-dessous (d'après Burton).

Logical order	Example
Numerical	a. 1939 b. 1940 c. 1941 d. 1942
Alphabetical	a. Changing *a* from .01 to .05. b. Decreasing the degrees of freedom. c. Increasing the spread of the exam scores. d. Reducing the size of the treatment effect.
Sequential	a. Heating ice from -100°C to 0°C. b. Melting ice at 0°C. c. Heating water from 0°C to 100°C. d. Evaporating water at 100°C. e. Heating steam from 100°C to 200°C.

Research. Numerous studies indicate that items are easier when this guideline is violated (Haladyna & Downing, 1989b).

La place de la réponse correcte devrait varier entre A-B-C et D. Des recherches ont cependant montré que les places B et C étaient les plus utilisées (McKeachie). Les étudiants aux faits de ces données vont choisir B ou C dans les cas où ils ne connaissent pas la bonne réponse, augmentant ainsi leur chance de réussite.



Analyser les réponses

Une fois le test passé, il est important d'effectuer une analyse des items du test pour déterminer l'efficacité du QCM. Il existe des logiciels de QCM disponibles sur internet qui permettent de calculer des statistiques pour chaque question en fonction de la difficulté de l'item, sa discrimination et la fréquence des réponses pour chaque option. Ce type d'analyse vous donne les informations nécessaires pour améliorer la validité et la fiabilité de vos QCM et estimer l'intensité des ressources (coût du temps de construction des items)(annexe 11)(Van der Vleuten, Schuwirth).

La difficulté d'un item est indiquée par le pourcentage d'étudiants qui ont répondu correctement appelé « facilité » de l'item. Etant donné que la chance de répondre correctement par hasard est de 25% pour 4 options, vous devez reformuler tout item qui atteint moins de 30%. Il est recommandé de chercher à écrire des items de niveaux de difficultés différents, avec une moyenne de difficulté d'environ 50%. Le but est de construire un test qui ne contient que peu d'items pour lesquels plus de 90% et moins de 30% des étudiants répondraient correctement (Davis). Les items sont considérés comme peu à modérément difficiles si 75 à 80% des étudiants répondent correctement. De plus, si vous formulez des questions que vous trouvez particulièrement faciles ou

difficiles et que les résultats ne correspondent pas à votre attente, vous devez retravailler vos QCM.

D'un autre point de vue, la validité de l'analyse des items existe par la bonne qualité de construction des items. Des études ont regardé l'effet des défauts de construction des items sur les résultats des étudiants. La plupart des études ont été faites sur 1 défaut (énoncé négatif, « aucune des autres options ») (Downing, Josefowicz). Downing a évalué la validité d'un test médical contenant des défauts. Il a trouvé que 33% (11/33) des items en contenait. Il est recommandé de faire lire et faire faire ses QCM à un collègue afin d'augmenter leurs qualités.

L'Indice de Discrimination d'un QCM se réfère à la différence de pourcentage des réponses correctes entre deux groupes d'étudiants (on se réfère en principe au top 25% et aux derniers 25%). Il sert à postériori à mesurer la qualité des QCM et à identifier les QCM à problèmes, c'est à dire à ID négatif (Damjanov). Il est défini par la formule suivante :

$$ID = \frac{1 - \frac{\%\ \text{réussite globale}}{\%\ \text{réussite des reçus}}}{\%\ \text{d'étudiants collés}}$$

Un ID positif signifie que les étudiants reçus ont mieux répondu que les autres et un ID négatif, que les étudiants reçus ont moins bien répondu que les autres. Dans ce dernier cas (discrimination négative), certains proposent d'éliminer la question de l'examen (Collins, CTL). Un item qui a une discrimination de 60% et plus est considéré comme très bon, alors qu'une discrimination inférieure à 19% indique que le QCM doit être révisé (Vydareny). Authier a montré que l'ID est meilleur lorsque l'énoncé est plus court et les propositions plus longues. Finalement, en regardant la fréquence des réponses incorrectes pour chaque question, on peut déterminer si ce sont des distracteurs équivalent. Si une réponse fausse n'a jamais été choisie, il est recommandé de réécrire le QCM (CTL).

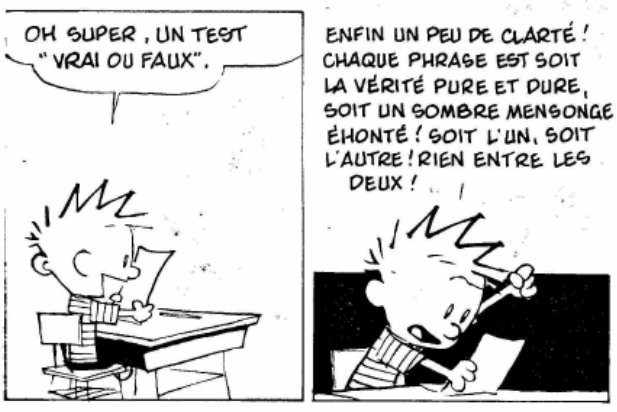
Conclusion

Les QCM sont un outil d'évaluation des connaissances objectif et rentable, largement utilisé dans l'enseignement supérieur. Ils ont leurs avantages et leurs limitations, comme tout outil d'évaluation.

Les critiques formulées par les enseignants et les étudiants viennent souvent d'un manque de connaissance des applications de cet outil d'évaluation et surtout de défauts de construction des items qui le constituent.

En effet, les enseignants ne sont pas tous familiers avec la rédaction d'objectifs éducationnels mesurables et de QCM correspondant à ces objectifs en fonction du niveau d'apprentissage à tester.

Leur rédaction doit suivre des règles de bonne pratique qui ont été rappelées dans ce mémoire. Lorsqu'ils sont bien formulés, les QCM peuvent mesurer de façon fiable, valide et reproductible les connaissances, la compréhension et les applications.


OH SUPER, UN TEST "VRAI OU FAUX".
ENFIN UN PEU DE CLARTÉ ! CHAQUE PHRASE EST SOIT LA VÉRITÉ PURE ET DURE, SOIT UN SOMBRE MENSONGE ÉHONTÉ ! SOIT L'UN, SOIT L'AUTRE ! RIEN ENTRE LES DEUX !

Références

- Authier. Calcul de l'indice de discrimination des QCM. Application à l'évaluation rétrospective de cinq épreuves d'embryologie et biologie du développement de PCEM1 à la faculté de médecine de Créteil/Paris XII. Mémoire de DIU de Pédagogie
- Bloom BS. Taxonomy of Educational Objectives: The Classification of Educational Goals; (Ed.) Susan Fauer Company, Inc. 1956: 201–207.
- Boland RJ, Lester NA, Williams E. Writing Multiple-Choice Questions. Academic Psychiatry. 2010; 34:310-316.
- Bonnefon D. Auteur de Questy http://www.questy.fr/
- Braddom CL. A brief guide to writing better test questions. Am J Phys Med Rehabil 1997 ; 76 : 514-516.
- Burton SJ, Sudweeks RR, Merrill PF, Wood B. How to prepare better Multiple-Choice Test Items : Guidelines for University Faculty. Brigham Young University Testing Service and The Department of Instructional Science. 1991.
- Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. Philapdelphia, PA : National Board of médical examiners, 1998.
- Cashin WE. Improving multiple-choice tests Idea paper Number 16. Manhattan, KS : Center for faculty development and évaluation, Kansas State University ; 1987.
- Collins J. Writing mutple choice questions for continuing médical éducation activities and self-assessment modules
- Cox KR, Bandaranayake R. How to write good multiple choice questions. Med J Aust 1978 ; 2 :553-554.
- Crooks TJ. The impact of classroom evaluation practices on students. Review of educational research. 1988, 58:438-481.
- CTL : Center for Teaching and Learning. University of North carolina at Chapel Hill. « For your consideration… » . Number 8 ; November 1990.
- Damjanov I, Fenderson B.A, Velovski J.J, Rubin E. Testing of médical students with open-ended, uncued questions. Human Pathol 1995 ;26 :362-5.
- Davis BG. Tools for teaching. Jossey-Bass Publishers. San Francisco, CA, chapter 30, 1993.
- Downing SM, Baranowski RA, Grosso LJ, Norcini JJ. Item type and cognitive ability measured : the validity évidence for multiple true-false items in médical speialty certification. Appl Meas Educ 1995 ; 8 :189-199.
- Eurich AC. Four types of examination. *J Educ Psychology* 1931 ; 22 :268-78.
- Gronlund NE . Assessment of student achievement. Boston, MA : Allyn and bacon, 1998.

- Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines. Applied Meas Educ 2002 ; 15 :309-333.
- Horgan J. Principes communs pour évaluer les résultats cognitifs de la formation. Service de technologie de l'éducation, Université de Liège, Belgique. Commission des communautés Européenne, 1991.
- Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew H. The quality of in-house médical school examinations. Acad Med 2002 ; 77 :156-161.
- Kemp JE, Morrison GR, Ross SM. Developing évaluation instruments. In : Designing Effective Instruction. New York, NY : macMillan College Publishing Company, 1994 :180-213.
- Kent TH, Jones JJ, Scmeiser CB. Some rules and guidelines for writing multiple choice test items. Iowa city : University of Iowa College of Medicine and American Testing Program. 1974.
- Leclercq D. La conception des QCM. Bruxelles : Labor. 1986.
- Lowman J. Mastering the techniques of teaching. San Francisco : Jossey-Bass, 1984.
- McCoubrie. Improving the fairness of multiple-choice questions : a littérature review. Medical teacher 2004 ; 26 :709-712.
- McKeachie WJ. Teaching Tips. 8th ed. Lexington, Mass. :Heath, 1986.
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. Med Educ 1985 ; 19 :238-247.
- Pagonis D. Guide de rédaction des questions à choix multiple. Université médicale virtuelle francophone.
- Sax G. Principles of educational measurement and évaluation. Belmont, CA : Wadsworth ; 1974.
- Srinivasa DK, Adkoll BV. Multiple choice questions : how to construct and how to evaluate ? Indian J pediatr 1989 ; 56 :69-74.
- Schulteis NM. Writing cognitive educational objectives and multiple-choice test questions. Am J Health-Syst Pharm 1998 ; 55 :2397-2401.
- Schuwirth LWT, van der Vleuten CPM. Different written assessment methods : what can be said about their strengths and weakness ? Medical Education 2004 ; 38 :974-979.
- Van der Vleuten CPM. The assessment of Professional compétence : developments, research and practical implications. Adv Health Scie duc 1996 ; 1 :41-67.
- Van Hoozer H et al. The teaching procès : theory and practice in nursing. Norwalk CT : Appleton-Century-Crofts, 1987 :279-280.
- Vydareny KH, Blane CE, Calhoun JG. Guidelines for writing multiple-choice questions in radiology courses. Invest Radiol 1986 ; 21 :871-876.
- Welsh AL. Multiple choice objective tests. In Saunders P, Welsh AL and Hansen WL (Eds.) ; Resource manual for teacher training programs in economics. New York : Joint Council on Economic Education ; 1978 : 191-228..

- Woodford K, Bancroft P. Using multiple choice questions effectively in information technology éducation. In R. Atkinson, C.MacBeath, D.Jonas-Dwyer and R.Philips (Eds), Beyond the comfort zone : Proceedings of the 21st ascilite Conference ; 2004 :948-955.
- Zimmaro DM. Writing Good Multiple-choice exams. Measurement and Evaluation center, Austin, 2004.

Tables et Annexes

Dans les annexes, les formulations incorrectes des exemples de QCM sont inscrites en italique alors que les bonnes formulations sont soulignées.

Table 1 : Exemples d'objectifs et QCM pour 3 niveaux différents de connaissances (d'après Collins)

Niveau	**Objectif**	**Item**
Connaissance (l'étudiant doit faire appel à sa mémoire)	Connaître la dose d'irradiation d'un scanner pulmonaire	Quelle est la dose d'irradiation moyenne d'un scanner pulmonaire ? A. 1mSv B. 8 mSv C. 16 mSv D. 24 mSv
Compréhension et Application (l'étudiant doit démontrer qu'il sait utiliser une information, appliquer des méthodes et des concepts à une autre situation)	Comparer les doses d'irradiation de différents examens radiologiques	Lequel des examens radiologique suivant est associé à la plus grande dose d'irradiation ? A. Scanner thoraco-abdominal B. Scanner des artères coronaires C. Radiographie pulmonaire conventionnelle D. Radiographie pulmonaire numérique
Analyse de problème (l'étudiant doit comprendre les données d'un concept, leurs relations entre elles et analyser des informations nouvelles en les scindant en différentes parties)	Expliquer les effets de différents facteurs sur les variations des doses d'irradiation d'un scanner pulmonaire	Laquelle des propositions suivantes diminue le moins la dose d'irradiation d'un scanner pulmonaire? A. diminuer les mA de 250 à 125 B. diminuer les kVp de 140 à 120 C. diminuer le temps du scanner de 2 à 1 D. diminuer le temps du scanner de 1 à 0.5

Table 2 : Guide de rédaction des QCM

- Items

  - Préparer un item de même niveau de connaissance que l'objectif à tester
  - Ecrire des items qui requièrent différents niveaux de connaissance
  - Dans l'énoncé et les options

  - Eviter les termes absolus « toujours », « jamais » « tous » « aucun »
  - Eviter les termes imprécis « parfois » « rarement » « généralement»« peu » « beaucoup »

- Enoncé

  - Poser un problème unique, complet et direct
  - Rédiger un énoncé bref, simple et précis, sans ambiguïté
  - Rédiger un énoncé sous forme de question et de forme positive
  - Eviter les indices permettant de découvrir la bonne réponse
  - Formuler des questions indépendantes (sans cascade ni tiroir)

- Options de réponse

  - Garder la même construction grammaticale que l'énoncé
  - N'exprimer qu'une seule notion par option
  - Préférer une seule bonne réponse par item
  - La bonne réponse doit être irréfutable
  - Garder une homogénéité de rédaction de la bonne réponse et des distracteurs en terme de style grammatical, longueur et complexité
  - Rédiger des distracteurs plausibles mais incontestablement faux
  - Eviter les termes « toutes les autres réponses » et « aucune des autres réponses »
  - Les options doivent être indépendantes et ne doivent pas se chevaucher
  - Ecrire les options en ordre logique (numérique, chronologique…) ce qui fait varier la position de la bonne réponse

Table 3 : Avantages et Inconvénients

▪ **Avantages**
o Permet d'évaluer un grand nombre d'étudiants o Versatilité: ils peuvent tester de nombreux sujets et à des niveaux d'objectifs éducationnels différents (Burton, Collins, CTL, Kemp, Pagonis, Woodford,) tels que tester les capacités de l'étudiant à : analyser un phénomène, appliquer un principe à de nouvelles situations, comprendre des concepts et des principes, discriminer les faits et les opinions, interpréter la relation cause-effet, interpréter des graphiques, juger la pertinence d'une information, émettre des déductions à partir de données, résoudre des problèmes. o Validité : représentatif de l'acquisition d'un grand nombre de connaissance (Burton, Collins, CTL, Pagonis, Zimmaro) o Fiabilité, Objectivité, Reproductibilité (Burton, Collins, CTL, Pagonis, Zimmaro) o Efficacité de correction, plus simple et rapide (Burton, Collins, CTL, Pagonis, Zimmaro) o Validation par une analyse des réponses
▪ **Inconvénients**
o Versatilité: étant donné que l'étudiant doit choisir une réponse à partir d'une liste d'alternatives plutôt que de construire lui même une réponse, les QCM ne sont pas adaptés à : organiser sa pensée, fournir des informations, produire des idées originales, donner des exemples (Burton) o Fiabilité. Les QCM sont exposés au risque de bonne réponse au hasard. On peut limiter ce facteur en augmentant le nombre d'items par test. Par exemple, si le test comporte 2 items de 4 propositions (A,B,C et D), on peut s'attendre à ce que 1/16 étudiant réponde correctement par hasard. Si la section comporte 15 items de 4 propositions chacun, on peut s'attendre à ce que 1/8,67 étudiant réponde correctement par hasard (Burton) o Peu valides si mal construits (Collins, Schuwirtz, Van der Vleuten)) o Difficiles et longs à construire, gestion lourde (1 heure par question) (Burton, Collins, Greco,Van Hoozer)

Annexe 1 : Est-ce que le QCM est adapté à ce que l'on veut tester ?

Pour s'avoir de façon fiable si un étudiant a atteint un objectif éducationnel, il est essentiel que la forme du test soit adaptée à l'objectif à tester. Les QCM sont souvent avantageux à utiliser mais pas dans toute les circonstances. C'est un outil particulièrement adapté pour évaluer les connaissances fondamentales (savoir). Les tests de concordance de script seront parfois plus adaptés pour évaluer le raisonnement, encore que (savoir comment), et l'observation directe pour évaluer le comportement (savoir être). En général, les QCM sont appropriés quand la réalisation d'un objectif d'éducation peut être mesuré par la sélection d'une réponse correcte parmi plusieurs alternatives (Burton). Par exemple :

- si la réalisation d'un objectif d'éducation peut être mieux mesuré par une réponse fournie par l'étudiant, un « QROC » ou une question rédactionnelle seront préférées
- si la réalisation d'un objectif d'éducation peut être mieux mesuré par une épreuve pratique, un test de performance doit être préféré

Les QCM sont des outils de recueil de l'information qui peuvent être utilisés pour évaluer les étudiants à plusieurs niveaux :

- Evaluation diagnostique (en amont d'une formation). Les QCM permettent d'estimer le niveau de connaissance des étudiants, ce qui permet d'adapter les cours.

- Evaluation formative (en cours de formation). Il est recommandé de fournir des solutions commentées pour chaque item, permettant à l'étudiant une auto-évaluation et une préparation aux examens.

- Evaluation sommative (en fin de formation). Il s'agit de valider les acquis cognitifs, méthodologiques et techniques en posant des questions centrées sur les objectifs qui ont fait l'objet d'un apprentissage préalable

Il a été proposé d'utiliser des QCM modifiés pour les intégrer aux Solutions Générales Implicites (SGI). On peut par exemple introduire des absurdités dan l'énoncé, ne fournir pas assez d'information, rédiger des questions où toutes les réponses sont vraies ou bien où toutes les réponses sont fausses. Bien sûr l'étudiant doit être averti de ces éventualités. Cela permet d'évaluer un niveau cognitif plus élevé, appelé la « vigilance cognitive » (Horgan, Leclercq).

Annexe 2 : Les différents types de QCM

1. Réponse Binaire (vrai/faux ; oui/non)

   Les hommes ont-ils un cerveau ?

   a. Oui

   b. non

2. QCM à choix simple (ou réponse unique)

   De quelle université fait partie l'hôpital Necker ?

   a. Paris 5

   b. Paris 6

   c. Paris 7

   d. Paris 11

3. QCM à choix libre (ou réponses multiples)

   Quels sont les CHU faisant partie de l'Université Paris-Descartes?

   a. Cochin

   b. Necker

   c. Pitié Salpétrière

   d. Sainte-Anne

4. Les appariements (ou associations)

   Quelles sont les services correspondant aux CHU suivants?

   1. Cardiologie adulte
   2. Neurochirurgie adulte
   3. Neurochirurgie Pédiatrique
   4. Psychiatrie adulte

   A. Cochin
   B. Necker
   C. Pitié Salpétrière
   D. Sainte-Anne

   a. 1D

   b. 2C

   c. 3B

   d. 4A

5. <u>Question illustrée (simple ou multiple, fait appel à une illustration, un schéma, un document sonore ou vidéo)</u>

   Quelle est la localisation de la lésion visible sur le scanner cérébral ci-dessous ?

   a. Angle Ponto-cerebelleux
   b. Hémisphère cérébral
   c. Tronc cérébral
   d. Vermis cérébelleux

Annexe 3 : Formuler des énoncés avec des phrases complètes

- *Question avec phrase à compléter*

Un hématome extra dural…

- <u>Question directe et complète</u>

<u>Laquelle des propositions suivantes caractérise un hématome extra dural au scanner ?</u>

Annexe 4 : L'énoncé doit contenir toute l'information utile et éviter de répéter la même phrase dans chaque option

- *Laquelle des options suivantes peut diminuer la dose d'irradiation par ½ ?*
  - *A. diviser 1 mA par 1/4*
  - *B. diviser 1 mA par 1/3*
  - *C. diviser 1 mA par 1/2*
  - *D. diviser 1 mA par 3/4*

- Par combien doit-on diviser un mA pour diminuer la dose d'irradiation de ½ ?
  - A. ¼
  - B. 1/3
  - C. ½
  - D. ¾

Annexe 5 : Les items peuvent mesurer différents niveaux de connaissance

- Item mesurant le rappel de connaissance (d'après Collins)

Laquelle des propositions suivantes représente une maladie chronique (plus de 3 mois) sur une radiographie pulmonaire ?

  A. Pneumonie à streptocoque
  B. Syndrôme de détresse respiratoire
  C. Œdème pulmonaire
  D. Protéinose alvéolaire pulmonaire

- Item mesurant l'application des connaissances (d'après Collins)

Un homme de 30 ans se présente avec une histoire de dyspnée, fébricule, toux et fatigue depuis 4 mois. Parmi les propositions suivantes quel est le diagnostic le plus probable?

  A. Pneumonie à streptocoque
  B. Syndrôme de détresse respiratoire
  C. Œdème pulmonaire
  D. Protéinose alvéolaire pulmonaire

Annexe 6 : Les énoncés doivent être formulés de façon positive

- *Eviter les formulations interro-négatives*

« Est ce que Sarkozy n'est pas le président ? »

- *Enoncé « négatif »*

Parmi les propositions suivantes, laquelle n'est pas une caractéristique de l'hématome extra dural au scanner ?

- Enoncé « positif »

Laquelle des caractéristiques radiologiques suivantes distingue un hématome extra dural d'un hématome sous-dural au scanner?

Annexe 7 : Formuler les distracteurs et la réponse correcte de la même longueur

- *L'option B est plus longue que les distracteurs*

  « Que signifie une déviation de la ligne médiane au scanner cérébral ? »

  A. Un déplacement de la tente

  B. Un déplacement de la faux par un processus expansif

  C. Un déplacement du corps calleux

  D. Un déplacement du tronc cérébral

- Toutes les options ont la même longueur

  « Que signifie une déviation de la ligne médiane au scanner cérébral ? »

  A. Un déplacement de la tente

  B. Un déplacement de la faux

  C. Un déplacement du corps calleux

  D. Un déplacement du tronc cérébral

Annexe 8 : Eviter les chevauchements d'options

- *Item avec chevauchement d'options*

  « Dans quelle gamme se situe la Pression IntraCrânienne normale? »

  A. 5-10 cm d'eau

  B. 10-15 cm d'eau

  C. 15-20 cm d'eau

  D. 20-25 cm d'eau

  - Item sans chevauchement d'options

  « Dans quelle gamme se situe la Pression IntraCrânienne normale? »

  A. 4-9 cm d'eau

  B. 10-15 cm d'eau

  C. 16-20 cm d'eau

  D. 21-26 cm d'eau

Annexe 9 : Eviter d'utiliser « toutes les autres réponses » et « aucune des autres réponses »

Comme il souligné par Burton, l'utilisation de ces termes peut diminuer la fiabilité d'un test, qu'ils constituent la bonne réponse ou les distracteurs, surtout quand on utilise « aucune des autres réponses ».

Alternative	Use	Weakness
"All of the above"	Answer	Can be identified by noting that two of the other alternatives are correct
	Distractor	Can be eliminated by noting that one of the other alternatives is incorrect
"None of the above"	Answer	Measures the ability to recognize incorrect answers rather than correct answers
	Distractor	Does not appear plausible to some students

Research. While research on the use of "all of the above" is not conclusive, the use of "none of the above" has been found in several studies to decrease item discrimination and test score reliability (Haladyna & Downing, 1989b).

Annexe 10 : Eviter de donner des indices de langage

- *L'option B ne suit pas la grammaire de l'énoncé*

  « Une déviation de la ligne médiane au scanner cérébral est _____ »

  A. Un déplacement de la tente

  B. Quand la faux est déplacée

  C. Un déplacement du corps calleux

  D. Un déplacement du tronc cérébral

- Toutes les options suivent la grammaire de l'énoncé

  « Une déviation de la ligne médiane au scanner cérébral est _____ »

  A. Un déplacement de la tente du cervelet

  B. Un déplacement de la faux

  C. Un déplacement du corps calleux

  D. Un déplacement du tronc créébral

Annexe 11 : Exemples de QCM informatisés

Ces liens (liste non exhaustive) vous permettent d'avoir accès à de la documentation sur l'écriture des QCM, des aides à l'élaboration de QCM en ligne et des outils de correction et d 'analyse des tests.

-AMC : Auto Multiple Choice. Correction automatisée de formulaires QCM. http://home.gna.org/auto-qcm/

-Dossiers thématiques de GRECO. http://greco.grenet.fr. Construction de tests d'évaluation et comparaison de 9 logiciels d'élaboration de QCM (points forts, points faibles). Propositions de différents systèmes de notation.

-Matexo fait partie de l'Anneau des mathématiques Francophones. http://matexo.smai.emath.fr/qcm/modemploi.php

-ModulEst. http://www.modulest.phedre.net/

-Moodle. http://e-uag.univ-ag.fr

-Netquiz Pro. http://www.ccdmd.qc.ca/

-Ordidac. http://dtup.free.fr/ordidac

- QCM mode d'emploi. www.irem.ups-tlse.fr/pim/act_tableur/**qcm**/**qcm_mode**_%20d'**emploi**.pdf

-QCM-QUIZZ PLUS

-Site de Bonnefon. Questy http://www.questy.fr/. Vous y trouverez en plus la description par des exemples des différents types de QCM

-Site de Rémi Bachelet. http://rb.ec-lille.fr/l/Correcteur_QCM.htm

-Site de Stéphane Metral. http://metral.info/qcm_en_ligne

-Site de Eric Dromer. http://gech.pagesperso-orange.fr/eval/qcm/frqcm.htm

-Thèbes. http://www.tpline.fr/telechargement/soft/Thebes.zip

S. Puget
DIU de pédagogie médicale
2010

QCM MODE D'EMPLOI

Résumé :

Les QCM sont maintenant largement utilisés dans l'enseignement supérieur bien que leur réputation auprès des enseignants et des étudiants soit mitigée. Ils sont en effet considérés par nature comme inférieurs aux questions ouvertes. Les étudiants qui font face à des QCM mal formulés les considèrent déroutant et réducteurs. Ces critiques à l'encontre des QCM sont plus souvent à attribuer à un défaut de construction des items qui le constituent plutôt qu'à une faiblesse intrinsèque.

Les enseignants ne sont pas tous familiers avec la rédaction d'objectifs éducationnels mesurables et de QCM correspondant à ces objectifs en fonction du niveau d'apprentissage à tester.

Nous proposons dans ce mémoire de résumer les bonnes pratiques de rédaction des QCM formulées dans la littérature et de proposer en fin de mémoire une liste des points clefs donnant une aide de rédaction des questions et des options pour rédiger des QCM. Il est primordial que les enseignants sachent rédiger des items valides afin d'être certain que les outils d'évaluation des connaissances utilisés sont valides. Les résultats de mesures des apprentissages sont utilisés pour construire de nouveaux buts d'apprentissage, ce qui complète le cycle continu de la connaissance.

Decouvrez Comment :

Rédiger les objectifs éducationnels

Définir les niveaux de connaissance pour chaque objectif

Rédiger des QCM appropriés pour tester ces connaissances

Analyser les résultats